

文章编号:1673-9469(2024)04-0098-05

DOI:10.3969/j.issn.1673-9469.2024.04.013

基于GPS轨迹数据的电动出租车充电站选址规划

任丹萍^{1,2},王茜茜^{1,2},陈湘国^{1,2*},邓玉静^{1,2}

(1. 河北工程大学 信息与电气工程学院,河北 邯郸 056038;

2. 河北工程大学 河北省安防信息感知与处理重点实验室,河北 邯郸 056038)

摘要: 针对电动汽车保有量持续增长、充电设施匮乏难以满足用户需求的问题,提出一种基于GPS轨迹数据的电动出租车充电站选址规划方案。首先利用出租车GPS数据分析用户潜在充电需求并提取需求分布;其次提出一种基于网格密度分区的DBSCAN聚类方法,与传统算法相比DB指数由0.34降为0.30,对需求进行聚类 and 划分需求密集区,设置预选站址;最后,构建集合覆盖模型实现站址优化。利用此方案对北京大兴区出租车轨迹数据进行仿真,得出了合理的选址结果,即该方案可为电动出租车充电站规划提供参考。

关键词: 充电站选址;电动出租车;GPS轨迹数据;密度聚类

中图分类号:TP399

文献标识码:A

Location Planning of Electric Taxi Charging Station Based on GPS Trajectory Data

REN Danping^{1,2}, WANG Xiqi^{1,2}, CHEN Xiangguo^{1,2*}, DENG Yujing^{1,2}

(1. School of Information and Electrical Engineering, Hebei University of Engineering, Handan, Hebei 056038, China;

2. Hebei Key Laboratory of Security Protection Information Sensing and Processing,
Hebei University of Engineering, Handan, Hebei 056038, China)

Abstract: A location planning scheme for electric taxi charging stations based on GPS trajectory data is proposed to address the problem of continuous growth in the number of electric vehicles and insufficient charging facilities to meet user needs. Firstly, the potential charging needs of users are analyzed using the taxi GPS data and demand distribution is extracted. Secondly, a DBSCAN clustering method based on the grid density partitioning is proposed, which reduces the DB index from 0.34 to 0.30 compared to traditional algorithms. The demand is clustered and divided into demand intensive areas, and pre-selected station locations are set. Finally, a set coverage model is developed to achieve site optimization. Using this scheme to simulate Daxing District, Beijing and taxi trajectory data, a reasonable site selection result is obtained, and the results can provide reference for the planning of electric taxi charging stations.

Key words: charging station location; electric taxi; GPS trajectory data; density clustering

近年来,新能源汽车产业发展迅猛,充电设施匮乏、分布不均等问题凸显,如何合理规划充电设施成为值得探索的课题。众多学者对此开展研究,其中以结合优化算法和影响因素构建选址模型为主流^[1-3],从建站方、用户的经济效益出发对充电站进行规划^[4-7],还有研究人员以提升用户充

电体验为目标对问题求解^[8]。由于电动出租车充电时间敏感度高、充电需求大、对公共充电站依赖性更强,研究人员更多挖掘轨迹数据进行选址研究^[9-10]。苏粟等挖掘乘客起讫点分布模拟出租车充电行为,综合时间成本规划充电站址^[11];邓昌棉等同样利用起讫点预估充电需求,以最小行驶距

收稿日期:2023-08-23

基金项目:河北省自然科学基金资助项目(F2022402007)

作者简介:任丹萍(1984-),女,河北邯郸人,博士,副教授,从事宽带通信网络方面的研究。

* 通信作者:陈湘国(1979-),男,河北邯郸人,硕士,讲师,从事物联网、计算机网络方面的研究。

离为目标构建选址模型并求解^[12]。上述两种基于乘客起讫点分布展开站点的选址研究,没有充分考虑建站成本。Keawthong 等分析轮班时出租车位置,以现存加油站数量作为潜在充电站数量,忽略了车辆运营期的充电需求和电动车与燃油车的补充能源时间差异^[13]。任峰等挖掘出租车行驶停靠点,结合区域负荷水平和车辆时序性完成选址^[14];罗思杰等聚类分析非道路区域出租车停靠点,选取停留密集区为充电站点^[15],这两种方法均对车辆停靠点进行聚类,而在实际运营中,无论车主停靠与否,充电行为发生在不同乘客乘车间隙,即车辆空载区,以保障出租车运营的经济效益。

综合当前电动出租车充电站选址存在的问题,考虑出租车充电特点和建站成本,提出一种基于出租车GPS轨迹数据的电动出租车充电站选址方法。首先对出租车轨迹数据进行分析,提取支持车辆完成充电的车辆空载期,将空载起点作为潜在预充电需求点;其次聚类挖掘需求密集区域,并将重心点设为充电站的预选址;最后构建集合覆盖模型,优化充电站布局。通过所选站址的需求覆盖分析对选址结果进行评价,验证方法的可行性。

1 选址方案

1.1 需求提取

进行城市公共设施布局时,轨迹数据的利用能帮助有效信息挖掘,反映交通特征^[16],因此本文基于出租车轨迹数据完成对充电站的选址分析。出租车轨迹数据记录信息为:车辆ID、采集时间、经度、纬度、载客状态等,载客状态由“0”和“1”表示,“0”表示当前无载客,“1”则有载。由于电动出租车主选取空载时段充电,因此提取支持充电时长的车辆空载期作为预充电需求,以起始位置作为预充电需求点。电动出租主要选用直流快充的方式进行充电,以我国主流电动出租车型比亚迪、上汽荣威等为例,快充时长主要在30~40 min之间^[15],故选取车辆空载40 min及以上为充电需求期,记为

$$T_{\text{dem}}^i = t_{10}^i - t_{11}^i \geq 40 \text{ min} \quad (1)$$

其中, T_{dem}^i 为车辆 i 的充电需求期, $l \in (0, 1)$ 表示车辆载客状态, t_{10}^i 为车辆 i 的空载起始时间点。车辆的预充电需求点可记为

$$P_{\text{dem}}^i = (p_{\text{on},10}^i, p_{\text{at},10}^i) \quad (2)$$

P_{dem}^i 为车辆 i 的预充电需求点, $p_{\text{on},10}^i$ 为车辆 i 的空载时刻经度, $p_{\text{at},10}^i$ 为车辆 i 的空载时刻纬度。提取充电需求的具体步骤如下:

算法 满足充电行为时长的充电需求提取算法
步骤1 输入出租车轨迹数据集 G 。

步骤2 原始数据清洗,对 G 进行去空、排序和降重:去除含有字段缺失的数据段;将原始数据先按照车辆ID进行升序排列,再按时间顺序进行升序排列;识别并删除重复数据段。

步骤3 清除非运营时间数据,保留早6:00至晚10:00的工作时间数据。

步骤4 对于相同车辆ID数据段,以时间升序检测数据段载客状态,若数据段 α 中载客状态为0, $\alpha+1$ 中状态为1,则删除数据段 $\alpha+1$,重复检测直至 $\alpha+i$ 中载客状态为1。

步骤5 计算 α 和 $\alpha+1$ 两点时间差 t ,若 $t < 40 \text{ min}$,则删除 $\alpha+1$ 及以上数据段,继续执行步骤4,否则执行步骤6。

步骤6 将数据段 $\alpha+1$ 以充电需求点保存,删除 $\alpha+1$ 及以上数据,转步骤4。

步骤7 结束。

1.2 需求分析

聚类作为轨迹数据的经典分析方法,广泛应用于商业选址等场景中。本文采用聚类方法分析充电需求密集区。由于轨迹数据密度分布不均,采用传统聚类方法分析效果不佳^[17],由此改进基于密度的噪声应用空间聚类(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)算法实现对不规则密度数据的聚类。基于网格密度分区的DBSCAN算法其主要思想为网格划分研究区域并区分密度等级分别进行聚类,提升聚类效果,具体实现过程为:依据研究范围构建矩形,对其进行行数和列数为 b 和 c 的网格划分,将网格裁剪至与研究区域相同形状;对网格和需求数据进行链接,计算不同网格的需求密度,使用自然间断点分级法 Jenks,将不同网格的密度值分为 n 个等级, Jenks 是一种依据数据分布规律分级分类的统计方法,该方法运用聚类思想,使落入同一级内部相似度最大,外部级与级间相异性最大,利用此方法可使同级区域范围内需求数据分布密度更均匀、相似度更高,如图1所示,相同颜色深度表示同密度等级区域。

对于同一密度值区域的充电需求点集 $P_{\text{dem}} = \{P_i\}$,若其中任意点 P_i 在给定半径 r 的邻域范围

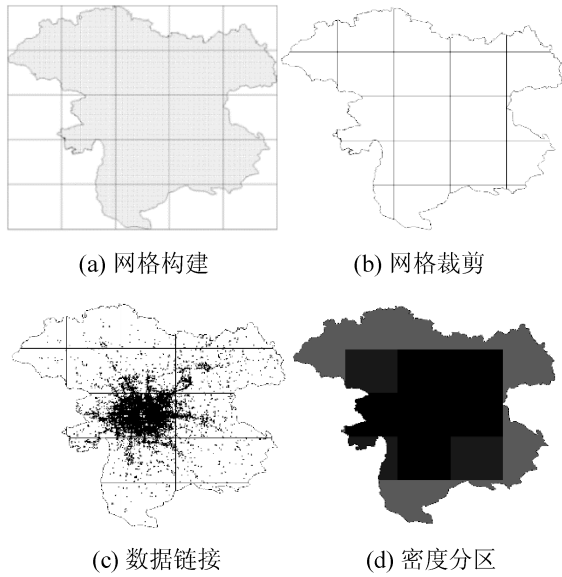


图1 网格密度分区

Fig. 1 Grid density partitions

内落入需求点数大于最小点数 MinPts , r 邻域内所有密度可达的点的集合为 D_i , 其中最大密度相连点集合为一簇 C , 表示为式(3), 其中 P_i 为当前簇的核心对象, haversine 为点 P_i 与 D_i 各点间的距离计算。聚类示意图如图2所示, 聚类簇之外的点为噪声点。

$$C(P_i) = \{D \in P_{\text{dem}} \mid \text{haversine}(P_i, D_i) \leq r\} \quad (3)$$

聚类操作完成后, 对于得出的每一簇 C 利用重心法求解重心点 N , 计算方法如式(4):

$$(P_{\text{on}}^N, P_{\text{at}}^N) = (\sum i \in cP_{\text{on}}^i / n, \sum i \in cP_{\text{at}}^i / n) \quad (4)$$

式中, N 则为该充电需求区的预选充电站点。

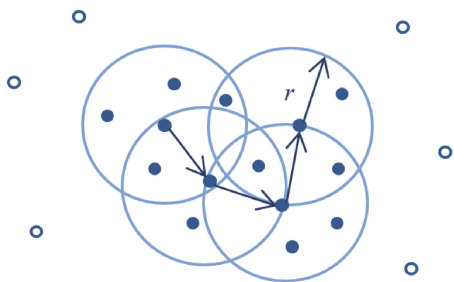


图2 DBSCAN 聚类示意图

Fig. 2 Schematic diagram of clustering by DBSCAN

1.3 集合覆盖模型构建

集合覆盖模型是以服务设施覆盖研究区域内所有需求为目标, 寻求设施的最小数量。考虑建站成本和地区充电服务需求, 利用集合覆盖模型完成站点的优化布局。

对于充电需求点集合 $P_{\text{dem}} = \{P_i \mid 1, 2, \dots, m\}$ 和预选站点集合 $N = \{N_j \mid 1, 2, \dots, n\}$ 构建集合覆盖模型, 目标函数为

$$\min F = \sum_{j=1}^n X_j \quad (5)$$

约束条件为

$$\sum_j \in M_i X_j \geq 1, i = 1, 2, \dots, m \quad (6)$$

$$X_j \in \{0, 1\}, j = 1, 2, \dots, n \quad (7)$$

式(5)表示满足需求覆盖的最小设施点数量, 约束(6)为确保任意充电需求点至少被一个充电站所覆盖, 其中 $M_i = \{j \mid d_{ij} = 1\}$ 表示覆盖需求点 i 的候选设施点的集合, d_{ij} 用于判断需求点是否具有服务覆盖, 若是则 $d_{ij} = 1$, 否则 $d_{ij} = 0$ 。约束(7)为决策变量 X_j 的取值范围。

2 实例分析

实验运行环境为 Windows10 操作系统, 硬件配置为 Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz 1.80 GHz 4 核 CPU, 内存 8 GB, 算法采用 Python 语言编写, 以北京市 2020 年 8 月 2 日至 8 日 10 357 辆出租车 GPS 轨迹数据为基础, 数据属性及示例如表 1 所示, 受硬件条件限制, 选取北京大兴区为实验区域进行分析。

表1 出租车轨迹数据属性

Tab. 1 Taxi track data properties

字段名称	属性	示例
ID	车辆 ID	4 568
Date	记录时刻	2020/8/5 10:12
JingDu	经度/(°)	116.358 018 4
WeiDu	纬度/(°)	39.932 319 18
Load	载客状态	1

2.1 数据处理

依据充电需求提取方法对数据进行处理, 得出充电需求数据 25 898 条。将需求数据和研究区路网数据可视化展示, 如图 3 所示。

由图 3 可见, 大兴新城及核心区和亦庄新城的大兴部分需求更密集, 向南、东南方向逐渐稀疏, 总体需求以西北方为中心呈放射状逐渐降低。

2.2 充电站选址

对潜在需求数据进行聚类挖掘需求密集区。对充电需求数据进行数据抽稀后得出日常可能存在的充电需求点 3 270 个, 对研究区域进行 3×3 网格划分和数据链接, 如图 4(a) 所示, 随着颜色加

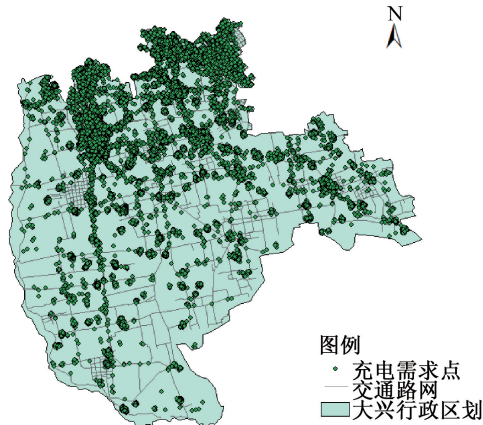


图 3 充电需求分布
Fig. 3 Charging demand distributions

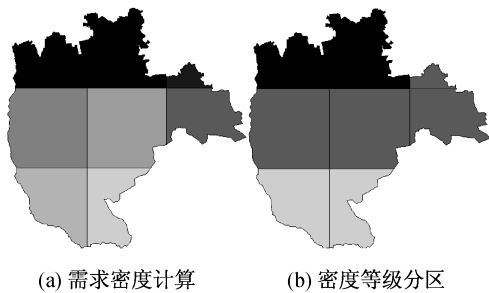


图 4 网格密度分区
Fig. 4 Grid density partitions

深,充电需求密度 ρ 逐渐增加。
采用自然间断点分级法对各研究区进行密度等级划分,结果如表 2,密度等级分区结果如图 4 (b)中所示。

表 2 网格密度等级划分 Tab. 2 Grid density classifications				
网格 序号	需求点 /pcs	面积 /km ²	需求密度 /(pcs·km ⁻²)	密度 等级
0	78	135. 340 604	0. 576 324	3
1	10	104. 272 29	0. 055 903	3
2	295	214. 150 114	1. 377 538	2
3	283	212. 228 602	1. 333 468	2
4	181	114. 628 294	1. 579 017	2
5	964	114. 811 296	8. 396 386	1
6	1 404	151. 388 941	9. 274 125	1
7	55	22. 846 244	2. 407 4	2

对相同等级区域内需求点进行 DBSCAN 聚类,聚类邻域半径 r 和最小样本数量 MinPts 通过计算需求数据的 K-平均最近邻距离形成候选参数组,当数据的分析结果趋于稳定时,选择 r 较大值时参数组作为最终参数^[18]。各密度等级区域进行分析结果见表 3,共得出充电需求密集区域

110 个。
求解各充电需求密集区需求重心,即得出预选充电站点,如图 5 所示。

表 3 不同密度等级区域聚类 r 和 MinPts Tab. 3 Regional clustering parameters r and MinPts under different density levels			
密度等级	邻域半径 r	最小样本量 MinPts	簇数
1	0. 001 748	8. 040 541	50
2	0. 001 248	3. 982 801	52
3	0. 001 146	2. 931 818	8

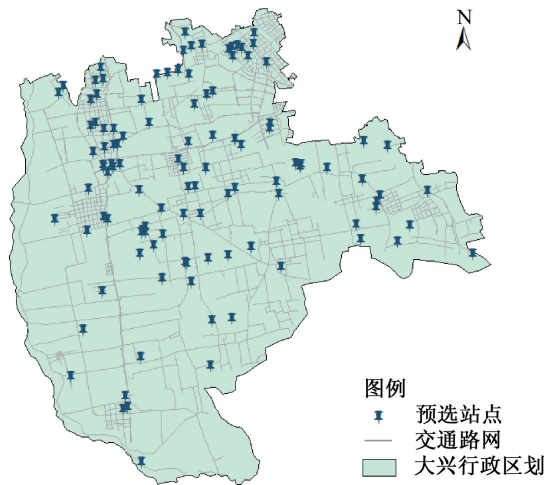


图 5 预选充电站点
Fig. 5 Pre-selected charging states

以充电需求点为服务请求,预选站点为待选设施点,根据北京市平原地区公共充电设施服务半径小于 3 km 的建设要求,设置中断阻抗为 3 km,构建集合覆盖模型并求解,得出优化后设施数量为 60 个,有效减少了设施建设数量。

3 结果分析

3.1 聚类效果分析
针对提出的基于网格密度分区的 DBSCAN 算法,引入 DB 指数 (Davies-Bouldin Index) 以评价聚类结果。DB 指数的运算为

$$R_{ij} = (s_i + s_j)/d_{ij} \tag{8}$$

$$DB = \frac{1}{N} \sum_{i,j=1}^N \max_{i \neq j} R_{ij} \tag{9}$$

式(8)为计算簇 i 与 j 间相似度,其中 s_i 表示簇 i 中所有样本点到簇中心距离均值, d_{ij} 为簇 i 与簇 j 中心点间距离, N 为簇的总数量。由式(9)可见,簇与簇间相似度越高、距离越近,DB 指数越高,聚类结果越差。分别对本文算法和传统算法的聚类

结果计算 DB 指数,本文算法为 0.30,传统算法指数为 0.34,可知本文方法的聚类结果质量更高,应用于对出租车充电需求的聚类更有优势。

3.2 选址结果分析

本文方法求得充电站点以服务半径为 3 km 的充电服务覆盖范围及服务范围内的充电需求量统计如图 6 所示。

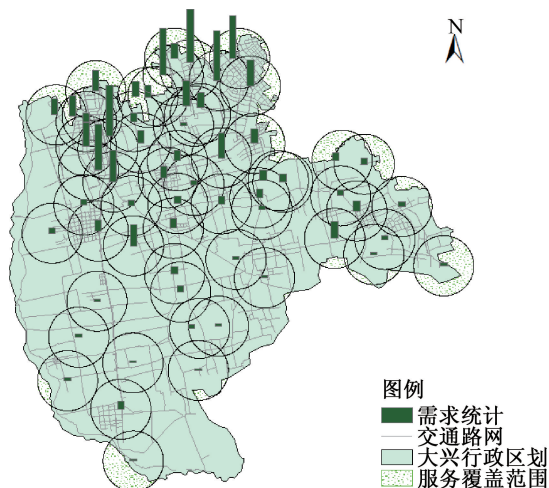


图 6 服务覆盖范围及数量

Fig. 6 Service coverage and quantity

不计服务重叠情况下,服务面积达 951.75 km²,优化前后对比如表 4 所示,由表可见,优化求解后在损失极小需求覆盖情况下,降低 45.5% 的设施规划成本,避免了建设浪费。同时与用未改进聚类算法进行选址的结果对比,更大的服务面积也印证了所提出聚类算法能识别分布更广的需求密集区,避免了传统算法面对不均匀密度分布数据时出现的噪声过多的问题。

表 4 设施数量及需求覆盖对比

Tab. 4 Comparison of facility quantity and demand coverage

聚类算法	构建集合覆盖模型	设施数量	需求覆盖率/%	覆盖面积/km ²
基于网格密度	是	60	98.93	951.75
分区的 DBSCAN	否	110	98.96	989.73
DBSCAN	是	52	94.15	706.61
	否	93	94.70	747.26

4 结论

1) 靠近城市的核心城区,出租车及行人的流量更大,电动出租车在运营期间的充电需求也会更高。

2) 本文提出的基于网格密度分区的 DBSCAN

算法对于密度分布不均的轨迹数据进行聚类时,相较于使用原始聚类算法,聚类效果有所提升。

3) 使用基于 GPS 轨迹数据的出租车充电站选址方法可有效覆盖区域充电需求,帮助车主降低充电时间和经济成本,计算结果可为电动出租车充电站选址提供参考。

参考文献:

- [1] WANG Y F, LIU D L, WU Y, et al. Locating and sizing of charging station based on neighborhood mutation immune clonal selection algorithm[J]. Electric Power Systems Research, 2023, 215(02): 109013.
- [2] 刘东林,王育飞,张宇,等. 基于 Huff 模型的电动汽车充电站选址定容方法[J]. 电力自动化设备, 2023, 43(11): 103-110.
- [3] ZHANG B, ZHAO M, HU X. Location planning of electric vehicle charging station with users' preferences and waiting time: Multi-objective bi-level programming model and HNSGA-II algorithm[J]. International Journal of Production Research, 2023, 61(05): 1394-1423.
- [4] ZHOU G Y, ZHU Z W, LUO S M. Location optimization of electric vehicle charging stations: Based on cost model and genetic algorithm[J]. Energy, 2022, 247(11): 123-134.
- [5] 李永竞,裴文卉. 基于多源数据的充电站三阶段选址优化决策[J]. 控制工程, 2023, 30(09): 1648-1657.
- [6] WANG L M, YANG C, ZHANG Y, et al. Research on multi-objective planning of electric vehicle charging stations considering the condition of urban traffic network[J]. Energy Reports, 2022, 8(11): 11825-11839.
- [7] 易校石,祁宝川,易正俊. 新能源电动汽车充电桩的优化选址[J]. 公路交通科技, 2022, 39(07): 166-171+180.
- [8] 梁迪,姜廷霖,郭启航. 基于改进鲸鱼优化算法的电动汽车充电设施选址[J]. 沈阳大学学报(自然科学版), 2022, 34(01): 24-29.
- [9] LI Y J, SU S, LUI B, et al. Trajectory-driven planning of electric taxi charging stations based on cumulative prospect theory[J]. Sustainable Cities and Society, 2022, 86(11): 104125.
- [10] YANG J, DONG J, HU L. A data-driven optimization-based approach for siting and sizing of electric taxi charging stations[J]. Transportation Research Part C: Emerging Technologies, 2017, 77: 462-477.
- [11] 苏粟,李玉璟,贾泽瑞,等. 基于 GPS 轨迹挖掘的电动出租车充电站规划[J]. 电力自动化设备, 2022, 42(10): 255-263.

(下转第 112 页)

- [16] SANDLER M, HOWARD A, ZHU M, et al. MobileNetv2: Inverted residuals and linear bottlenecks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
- [17] DING X, ZHANG X, MA N, et al. Repvgg: Making vgg-style convnets great again [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.
- [18] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
- [19] 樊嵘, 马小陆. 面向拥挤行人检测的改进 DETR 算法 [J]. 计算机工程与应用, 2023, 59(19): 159-165.
- [20] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021 [J]. arXiv preprint arXiv:2107.08430, 2021.
- (责任编辑 王利君)

(上接第 102 页)

- [12] 邓昌棉, 张勇. 基于数据驱动的电动出租车充电站规划方法研究 [J]. 森林工程, 2020, 36(03): 77-85.
- [13] KEAWTHONG P, MUANGSIN V, GOWANIT C. Location selection of charging stations for electric taxis: A bangkok case [J]. Sustainability, 2022, 14(17): 11033.
- [14] 任峰, 向月, 雷小林, 等. 基于 GPS 数据的电动出租车充电桩选址定容 [J]. 电力自动化设备, 2022, 42(10): 273-279.
- [15] 罗思杰, 邹复民, 郭峰, 等. 基于轨迹数据的出租车充电站选址方法 [J]. 计算机工程与应用, 2022, 58(08): 273-282.
- [16] 张龙深, 葛星, 刘迪. 基于互联网开放数据的城市实时交通圈提取及分析 [J]. 河北工程大学学报(自然科学版), 2017, 34(04): 94-98.
- [17] 任丹萍, 刘琳, 陈湘国. 基于密度分区的出租车载客热点区域聚类分析 [J]. 计算机应用与软件, 2023, 40(10): 83-89.
- [18] 李文杰, 闫世强, 蒋莹, 等. 自适应确定 DBSCAN 算法参数的算法研究 [J]. 计算机工程与应用, 2019, 55(05): 1-7+148.
- (责任编辑 王利君)

《河北工程大学学报(自然科学版)》2024 年起由季刊变更为双月刊

为进一步缩短论文出版周期, 更快推动科研成果的传播, 满足广大专家学者学术交流需要, 经主管主办单位同意, 申请期刊刊期变更, 获得河北省新闻出版局批复(冀新闻出版[2024]-060000001), 《河北工程大学学报(自然科学版)》于 2024 年第 1 期起由季刊变更为双月刊, 每逢双月的 25 号出版, 特此公告。